

What is claimed is:

1. A method for processing query messages over a network, comprising:
 - extracting a plurality of queries from a plurality of query messages received from a plurality of users over the network;
 - determining a number of queries included in the plurality of queries;
 - associating a current sequence number with the plurality of queries;
 - creating a request message including the plurality of queries, a first sequence number equal to the current sequence number and a first message count equal to the number of queries;
 - sending the request message to a search engine;
 - receiving a response message from the search engine, the response message including a plurality of replies, a second sequence number, a second message count, a third sequence number and a third message count;
 - creating a plurality of reply messages from the plurality of replies; and
 - sending the plurality of reply messages to the plurality of users over the network.
2. The method of claim 1, wherein:
 - the second sequence number equals the first sequence number; and
 - the second message count is equal to or less than the first message count.
3. The method of claim 2, wherein:
 - the third sequence number does not equal the first sequence number; and
 - the third message count is greater than zero.
4. The method of claim 1, further comprising:
 - determining a message latency associated with the first sequence number.
5. The method of claim 4, wherein said determining a message latency includes:
 - updating a request timestamp based on the request message;
 - updating a response timestamp based on the response message; and

comparing the request timestamp and the response timestamp.

6. The method of claim 5, further comprising:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and

updating the response timestamp based on the additional response message.

7. The method of claim 4, wherein said determining a message latency includes:

updating a query count based on the request message;

updating a reply count based on the response message; and

comparing the query count and the reply count.

8. The method of claim 7, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and

updating the reply count based on the additional response message.

9. The method of claim 4, wherein said determining a message latency includes:

updating a response count based on the response message; and

comparing the response count to a predetermined response count.

10. The method of claim 9, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and

updating a response count based on the additional response message.

11. A system for processing query messages over a network, comprising:

 - a first network interface coupled to a first network;
 - a second network interface coupled to a second network;
 - at least one processor coupled to the first network interface and the second network interface; and
 - a memory coupled to the processor, the memory including instructions adapted to be executed by the processor to:

 - extract a plurality of queries from a plurality of query messages received from a plurality of users over the first network interface;
 - determine a number of queries included in the plurality of queries;
 - associate a current sequence number with the plurality of queries;
 - create a request message including the plurality of queries, a first sequence number equal to the current sequence number and a first message count first message count equal to the number of queries;
 - send the request message to a search engine over the second network interface;
 - receive a response message from the search engine over the second network interface, the response message including a plurality of replies, a second sequence number, a second message count, a third sequence number, and a third message count;
 - create a plurality of reply messages from the plurality of replies; and
 - send the plurality of reply messages to the plurality of users over the first network interface.
12. The system of claim 11, wherein the first network and the second network are the same network.
13. The system of claim 11, wherein:

 - the second sequence number equals the first sequence number; and
 - the second message count is equal to or less than the first message count.

14. The system of claim 13, wherein:

the third sequence number does not equal the first sequence number; and
the third message count is greater than zero.

15. The system of claim 11, wherein the instructions are further adapted to:

determine a message latency associated with the first sequence number,
including:

update a request timestamp based on the request message,

update a response timestamp based on the response message, and

compare the request timestamp and the response timestamp;

receive an additional response message from the search engine, the additional
response message including an additional plurality of replies, a fourth sequence number
equal to the first sequence number and a fourth message count greater than zero; and

update the request timestamp based on the additional response message;

16. The system of claim 11, wherein the instructions are further adapted to:

determine a message latency associated with the first sequence number,
including:

update a query count based on the request message,

update a reply count based on the response message, and

compare the query count and the reply count;

receive an additional response message from the search engine, the additional
response message including an additional plurality of replies, a fourth sequence number
equal to the first sequence number and a fourth message count greater than zero; and

update the reply count based on the additional response message.

17. The system of claim 11, wherein the instructions are further adapted to:

determine a message latency associated with the first sequence number,
including:

update a response count based on the response message, and

compare the response count to a predetermined response count;

receive an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and update a response count based on the additional response message.

18. A computer readable medium including instructions adapted to be executed by at least one processor to implement a method for processing query messages over a network, the method comprising:

extracting a plurality of queries from a plurality of query messages received from a plurality of users over the network;

determining a number of queries included in the plurality of queries;

associating a current sequence number with the plurality of queries;

creating a request message including the plurality of queries, a first sequence number equal to the current sequence number and a first message count equal to the number of queries;

sending the request message to a search engine;

receiving a response message from the search engine, the response message including a plurality of replies, a second sequence number, a second message count, a third sequence number and a third message count;

creating a plurality of reply messages from the plurality of replies; and

sending the plurality of reply messages to the plurality of users over the network.

19. The computer readable medium of claim 18, wherein:

the second sequence number equals the first sequence number; and

the second message count is equal to or less than the first message count.

20. The computer readable medium of claim 19, wherein:

the third sequence number does not equal the first sequence number; and

the third message count is greater than zero.

21. The computer readable medium of claim 18, wherein the method further comprises:

determining a message latency associated with the first sequence number.

22. The computer readable medium of claim 21, wherein said determining a message latency includes:

updating a request timestamp based on the request message;

updating a response timestamp based on the response message; and

comparing the request timestamp and the response timestamp.

23. The computer readable medium of claim 22, wherein the method further comprises:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and

updating the request timestamp based on the additional request message;

24. The computer readable medium of claim 21, wherein said determining a message latency includes:

updating a query count based on the request message;

updating a reply count based on the response message; and

comparing the query count and the reply count.

25. The computer readable medium of claim 24, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and

updating the reply count based on the additional response message.

26. The computer readable medium of claim 21, wherein said determining a message latency includes:

updating a response count based on the response message.

27. The computer readable medium of claim 26, wherein said determining a message latency includes:

receiving an additional response message from the search engine, the additional response message including an additional plurality of replies, a fourth sequence number equal to the first sequence number and a fourth message count greater than zero; and
updating a response count based on the additional response message.